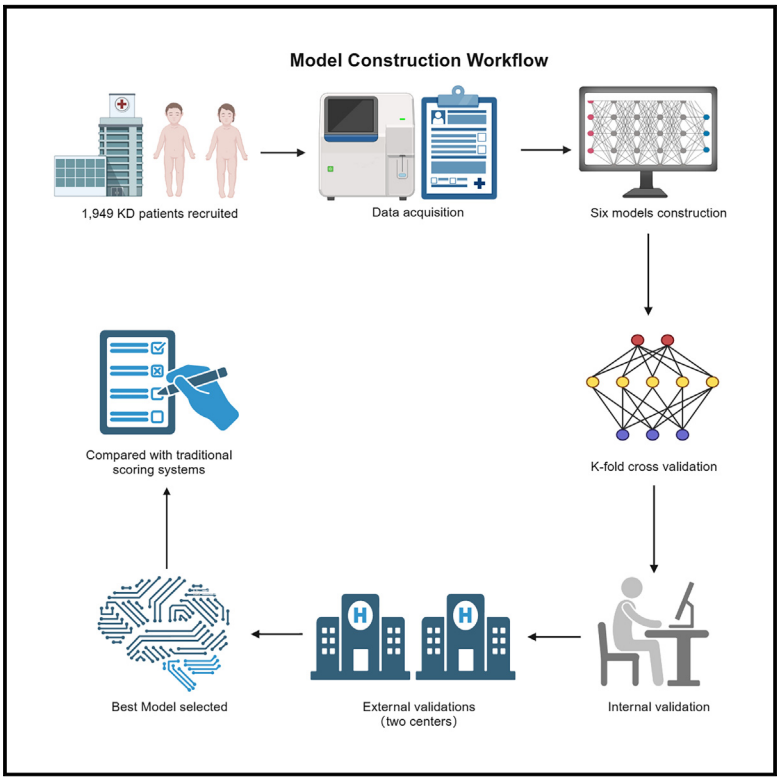
iScience

Article

A machine learning-based model to predict intravenous immunoglobulin resistance in Kawasaki disease

Graphical abstract



Authors

Yuhan Xia, Yuezhong Huang, Min Gong, ..., Rongzhou Wu, Maoping Chu, Xiu-Feng Huang

Correspondence

chmping@hotmail.com (M.C.), hxfwzmc@163.com (X.-F.H.)

In brief

Cardiovascular medicine; Machine learning

Highlights

- We developed predictive models for IVIG resistance in Kawasaki disease
- Our model outperformed other models, with AUCs ranging from 0.711 to 0.827
- The model underwent internal and external validation, confirming its robustness
- Key predictors include CRP-to-albumin ratio, prognostic index, and sex

Xia et al., 2025, iScience *28*, 112004
March 21, 2025
https://doi.org/10.1016/j.isci.2025.112004

CellPress

iScience

CellPress
OPEN ACCESS

Article

A machine learning-based model to predict intravenous immunoglobulin resistance in Kawasaki disease

Yuhan Xia,[1,2,3,6] Yuezhong Huang,[1,2,6] Min Gong,[4,6] Weirong Liu,[5] Yuanhui Meng,[3] Huiyang Wu,[3] Hui Zhang,[3] Hao Zhang,[3] Luyi Weng,[3] Xiao-Li Chen,[1,2] Huixian Qiu,[3] Xing Rong,[3] Rongzhou Wu,[3] Maoping Chu,[1,2,3,*] and Xiu-Feng Huang[1,2,7,*]

[1]Zhejiang Provincial Clinical Research Center for Pediatric Precision Medicine, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China

[2]Pediatrics Discipline Group, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China

[3]Children's Heart Center, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, China

[4]Department of Pediatrics, People's Hospital of Quzhou, Quzhou, Zhejiang, China

[5]Department of Pediatrics, People's Hospital of Shaoxing, Shaoxing, China

[6]These authors contributed equally

[7]Lead contact

*Correspondence: chmping@hotmail.com (M.C.), hxfwzmc@163.com (X.-F.H.)

https://doi.org/10.1016/j.isci.2025.112004

SUMMARY

Accurate prediction of intravenous immunoglobulin (IVIG) resistance is crucial for the effective treatment of Kawasaki disease(KD). This study aimed to develop a predictive model for IVIG resistance in patients with Kawasaki disease and to identify the key predictors. The training set underwent cross-validation, and models were constructed using six machine learning algorithms. Model performance was validated through cross-validation, test set evaluation, and two external validation sets evaluation. The model constructed using the random forest algorithm demonstrated the best overall performance among six models. The areas under the receiver operating characteristic curve (AUCs) for 5-fold cross-validation, internal validation, and external validations from Shaoxing and Quzhou were 0.711, 0.751, 0.827, and 0.735, respectively. According to the Shapley additive explanation (SHAP) method, C-reactive protein-to-albumin ratio, prognostic nutritional index, and sex were identified as the most important predictors. Our model demonstrates strong predictive capability for assessing IVIG resistance in Kawasaki disease patients.

INTRODUCTION

Kawasaki disease (KD) is an acute systemic vasculitis in young children that affects medium-sized arteries, specifically the coronary arteries.[1] Coronary artery lesions (CALs), mainly comprising coronary artery dilatations and coronary artery aneurysms, stand as the predominant complications of KD.[2,3] In developed countries, KD has emerged as a leading cause of acquired heart disease in children. High-dose intravenous immunoglobulin (IVIG) treatment (2 g/kg) has been shown to effectively reduce the incidence of CALs. However, approximately 10%–20% of affected patients exhibit resistance to IVIG treatment.[4,5] Extensive evidence highlights a strong correlation between IVIG resistance and the occurrence of CAL.[6–8] Recent studies indicate that administering corticosteroids in conjunction with standard-dose IVIG as an initial treatment for children at high risk of IVIG resistance can significantly decrease the incidence of coronary artery abnormalities, thereby alleviating complications associated with Kawasaki disease.[9] Consequently, accurately identifying KD patients at high risk of IVIG resistance is crucial for implementing early therapeutic strategies, essential for preventing the development of CALs in these individuals.

Previously, Egami,[10] Kobayashi,[11] and Sano[12] developed three distinct scoring systems for IVIG resistance based on characteristics observed in the Japanese population. However, their efficacy is largely restricted to the specific populations studied. Their applicability to other populations, particularly non-Japanese cohorts, is limited, likely due to demographic variability. As a result, various regions have established regionally specific IVIG resistance scoring systems. Lin et al. based on the variables included albumin, percentage of neutrophils, and positive lymphadenopathy to establish a simple three-variable score to predict the IVIG unresponsiveness for patients in Taiwan.[13] Yang et al. construct a predictive tool for the efficacy of IVIG therapy in children with KD in Beijing.[14] Although both models were developed primarily based on Han Chinese populations, their predictive performance was poor when applied to the Chongqing region, which is also predominantly the same population. This difference in predictive performance is likely not solely attributed to ethnic variations but may also be caused by the potential flaws in the model construction methodology,[15] as the aforementioned predictive models were

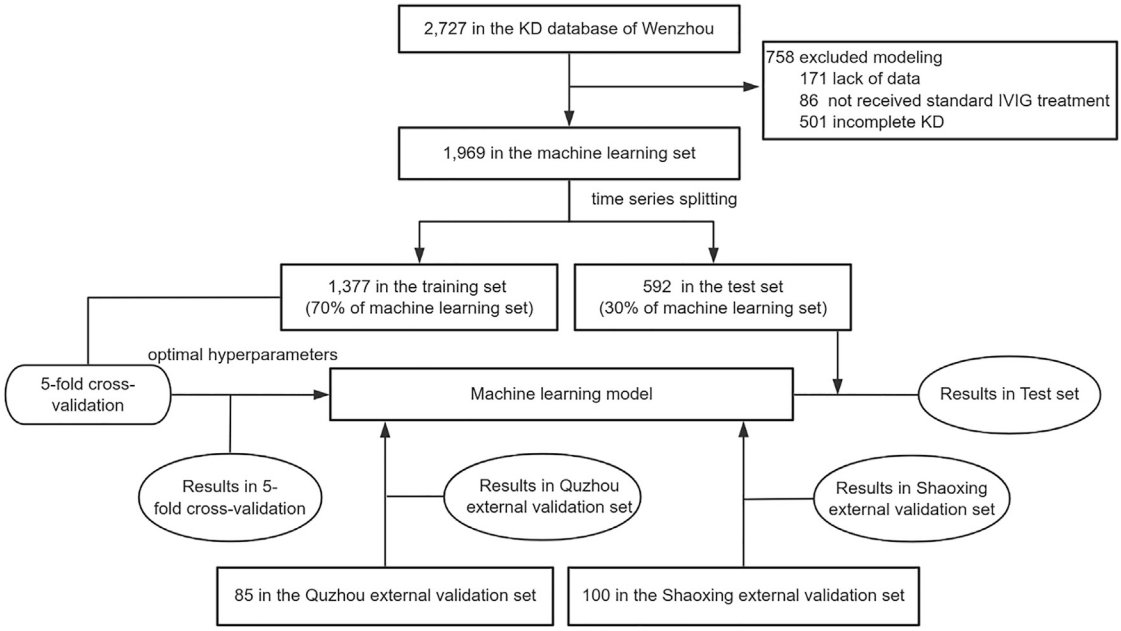


Figure 1. Study design
Abbreviations: KD, Kawasaki disease; IVIG, intravenous immunoglobulin.

developed using logistic regression algorithms. Nevertheless, traditional predictive models based on logistic regression often face significant limitations, particularly when addressing nonlinear relationships between input features and outcomes, or when accounting for interactions among features.

One notable advantage of machine learning approaches is their capacity to capture and analyze complex, nonlinear relationships among variables, resulting in more consistent and reliable predictions. Over the past decade, these algorithms have been increasingly applied across diverse disciplines, gaining particular traction in healthcare. Their ability to predict patient outcomes with greater accuracy than conventional methods has been demonstrated in numerous clinical contexts and disease states, establishing them as valuable tools in medical research and practice. Recently, machine learning has also shown considerable promise in assisting the diagnosis of KD.[16,17] Herein, this study aims to develop a machine learning-based model for KD, and identify key predictors of IVIG resistance based on the established model.

RESULTS

Demographic and clinical characteristics of patients

As shown in the flowchart of Figure 1, a total of 1,949 children, including 244 cases of IVIG resistance (12.4%), were identified from Wenzhou Hospital's database and included in the machine learning set. These children had complete data, received standard IVIG treatment, and fulfilled the diagnostic criteria for complete KD. The median age was 22.0 months (interquartile range: 12.0–38.0 months), 1,238 cases (62.8%) were male, and 244 cases (12.4%) were resistant to IVIG. The dataset was divided into training and test sets in a 7:3 ratio according to time series. Thus, 1,377 patients (160 with IVIG resistance, 11.6%) hospitalized between January 2008 and August 2019 were assigned to the training set, while 592 patients (84 with IVIG resistance, 14.2%) admitted between September 2019 and October 2022 were used as the test set. Demographic and laboratory characteristics were compared between the training and test sets to assess baseline distribution (Table S1). The remaining 758 children were excluded due to failure to meet the aforementioned criteria and were designated as the excluded set. In order to assess the impact on the study of the excluded set, we compared the demographic and laboratory characteristics of the excluded set and the machine learning set. The results showed no difference in IVIG resistance rates between the two sets ($p = 0.105$) (Table S2).

For external validation, 100 KD patients (5 with IVIG resistance, 5.0%) from Shaoxing Hospital from March 2020 to June 2024 and 85 KD patients (5 with IVIG resistance, 5.9%) from Quzhou Hospital from February 2017 to October 2023 were included. These are hereinafter referred to as the Shaoxing external set and the Quzhou external set, respectively. We conducted pairwise comparisons of patient characteristics across the three centers based on the selected features to identify potential differences in the patient populations (Table S3). The IVIG resistance rate in Wenzhou Hospital was numerically higher (12.3%) than that of Shaoxing Hospital (5.0%) and Quzhou Hospital (5.9%). However, statistical analysis revealed a significant difference only between Wenzhou and Shaoxing Hospitals ($p = 0.027$), while no significant difference was observed between Wenzhou and Quzhou Hospitals ($p = 0.072$).

We also compared the clinical features of IVIG-responsive and IVIG-resistant patients across the three centers (Table S4) to evaluate differences in treatment response. Consistent with expectations, IVIG-resistant patients exhibited elevated markers

Table 1. Comparison of six machine learning models in the 5-fold cross-validation

Model	AUC[a]	Sensitivity[a]	Specificity[a]	Accuracy[a]
Random forest	0.711 (0.642–0.790)	0.631 (0.562–0.781)	0.713 (0.504–0.901)	0.703 (0.536–0.862)
XGBoost	0.704 (0.640–0.747)	0.713 (0.625–0.875)	0.612 (0.481–0.749)	0.624 (0.513–0.735)
Logistic regression	0.692 (0.611–0.796)	0.613 (0.500–0.688)	0.708 (0.605–0.831)	0.697 (0.611–0.796)
LGBM	0.683 (0.654–0.732)	0.550 (0.469–0.688)	0.779 (0.683–0.824)	0.752 (0.684–0.793)
SVM	0.662 (0.630–0.727)	0.575 (0.375–0.781)	0.728 (0.535–0.926)	0.710 (0.564–0.865)
CatBoost	0.704 (0.627–0.749)	0.694 (0.594–0.812)	0.677 (0.543–0.782)	0.679 (0.564–0.767)

AUC, area under the receiver operating characteristic curve; XGBoost, eXtreme gradient boosting; LGBM, light gradient boosting machine; SVM, support vector machine; CatBoost, categorical boosting.
[a]Range of 5-fold cross-validation results is indicated in parenthesis.

associated with inflammation, while albumin (ALB), sodium (NA), and prognostic nutritional index (PNI) levels, among other markers negatively correlated with inflammation, were lower, reflecting a more severe inflammatory state. The most pronounced differences were observed at Wenzhou Hospital, where 18 of 22 variables showed significant differences between IVIG-responsive and IVIG-resistant patients.

Model construction and evaluation

We constructed and validated the machine learning models as described in the STAR Methods. The optimal hyperparameters for each model during the 5-fold cross-validation are detailed in Table S5. The results of each model's performance during cross-validation are recorded in Table 1. Among the six models, the random forest model demonstrated the best performance with an average area under the receiver operating characteristic curve (AUC) of 0.711 (range: 0.642–0.790), followed by the XGBoost (AUC = 0.704, range: 0.640–0.747) and CatBoost models (AUC = 0.704, range: 0.627–0.749). The average AUCs of the remaining models were below 0.700. The random forest model achieved average sensitivity, specificity, and accuracy of 0.631 (range: 0.562–0.781), 0.713 (range: 0.504–0.901), and 0.703 (range: 0.536–0.862), respectively.

To evaluate the generalizability of the models, the test set and two external validation sets were employed. The ROC curves for the six machine learning models on the test set are shown in Figure 2. In the test set validation, the random forest model outperformed the others, achieving an AUC of 0.751 (95% confidence interval [CI]: 0.692–0.820), with a sensitivity of 0.726 (95% CI: 0.631–0.821), specificity of 0.720 (95% CI: 0.681–0.764), and accuracy of 0.721 (95% CI: 0.686–0.758) (Table 2). In the Shaoxing external validation set, the light gradient boosting machine (LGBM) model showed the highest AUC of 0.893 (95% CI: 0.808–0.968), followed by the random forest model with an AUC of 0.827 (95% CI: 0.659–0.949). Similarly, in the Quzhou external validation set, the LGBM model again demonstrated the highest AUC at 0.743 (95% CI: 0.540–0.888), with the random forest model ranking second with an AUC of 0.735 (95% CI: 0.595–0.860). DeLong analysis indicated no statistically significant differences between these two models in the internal validation ($p = 0.456$), Shaoxing validation ($p = 0.511$), or Quzhou validation ($p = 0.940$). Due to the inherent dependence of samples in 5-fold cross-validation, DeLong analysis was not performed in this context. Although the DeLong test showed no significant differences between the models, we selected the random forest model for SHAP analysis and sample size assessment due to its robust and consistent performance across 5-fold cross-validation, internal validation, and external validation.

SHAP features

The SHAP analysis results are presented in Figure 3. In the test set validation, the top three SHAP features were the C-reactive protein-to-albumin ratio (CAR, SHAP value = 0.038), PNI (0.029), and sex (0.028). In the Quzhou external validation, the top three SHAP features were also CAR (0.033), sex (0.030), and PNI (0.029). In the Shaoxing external validation, CAR (0.032) and sex (0.028) were the top two features, with PNI (0.025) ranking fourth. Given the important role of sex in KD, we performed a stratified analysis to explore its impact further (Figure S1). The results show that CAR and PNI consistently ranked as the two most influential predictors and the ranking of the other indicators also did not differ significantly between men and women.

Sample size assessment

We conducted a learning curve analysis for the random forest model to assess whether the sample size used to construct the model was adequate (Figure S2). The results showed that the AUC increased significantly with the initial expansion of the training set, up to 600 samples. Beyond this point, the AUC exhibited minor fluctuations, initially decreasing slightly before stabilizing. Notably, the difference in AUC between the 600-sample training set and the final training set size was minimal, indicating that the model's performance had effectively stabilized.

Comparison with the previous predictive models

To determine whether the constructed random forest model offers an advantage over traditional scoring systems, we compared its performance on the validation set with previous models. The confusion matrices for each model are shown in Figure S3. The random forest model (AUC = 0.751, 95% CI: 0.692–0.820) outperformed the Egami[10] (AUC = 0.650), Kobayashi[11] (AUC = 0.646), Sano[12] (AUC = 0.658), Formosa[13] (AUC = 0.590), and Fu[14] (AUC = 0.621) models (Table 3).

DISCUSSION

In this study, we conducted a multicenter investigation involving 2,154 patients with KD across three distinct centers to develop a

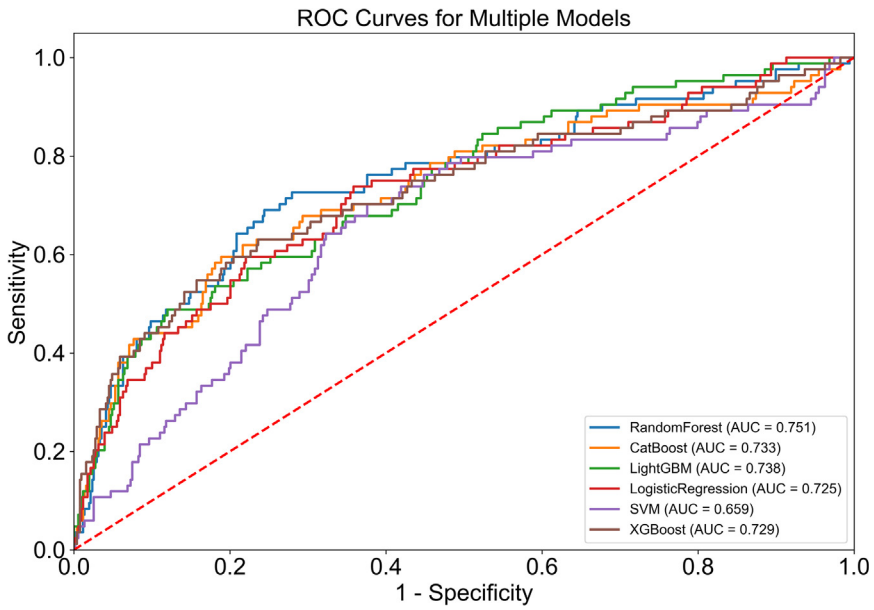


Figure 2. ROC curves of six machine learning models in the test set

The horizontal axis indicates false-positive rate (1 -specificity), and the vertical axis indicates true positive rate (sensitivity). AUC is indicated at the bottom. Abbreviations: AUC, area under the receiver operating characteristic curve; XGBoost, eXtreme gradient boosting; LGBM, light gradient boosting machine; SVM, support vector machine; CatBoost, categorical boosting.

machine learning-based model for predicting IVIG resistance. The key findings of this study were as follows: (1) the machine learning model, developed using a random forest algorithm, demonstrated robust performance in identifying IVIG-resistant patients among those with KD; and (2) CAR, PNI, and sex were identified as critical predictors of IVIG resistance.

Historically, the prediction of IVIG resistance has relied on traditional scoring systems based on a fixed set of predefined variables. However, these conventional methods are often limited by multicollinearity among variables, which can reduce the accuracy and reliability of predictions. In contrast, machine learning algorithms, with their ability to handle complex interactions and select relevant features, offer a significant advantage by effectively managing multicollinearity and capturing more nuanced patterns that may be missed by traditional models. In this study, we compared our machine learning model to previously established IVIG resistance scoring systems[10–14] and found that it exhibited significantly superior predictive capability. This enhanced performance was consistent even when compared to the Beijing scoring system, specifically designed for Chinese populations, suggesting that the model's superiority is not due to population-based differences.

Moreover, we observed that patients from Wenzhou Hospital exhibited a higher rate of IVIG resistance and elevated levels of certain inflammatory markers, such as white blood cell count (WBC), absolute neutrophil count (ANC), and systemic immune-inflammatory index (SII). This may be attributable to Wenzhou Hospital's role as a specialized pediatric center, where more severe cases are referred. Despite these differences in patient characteristics, the machine learning model trained on the Wenzhou cohort demonstrated strong predictive performance in external validation sets, further supporting its utility in diverse clinical settings.

We also compared our model to several existing IVIG resistance predictive machine learning models.[18–20] Among these, only Kuniyoshi et al. employed K-fold cross-validation similar to our study. Our model significantly outperformed their full-variable model (average AUC = 0.711, range: 0.642–0.790 vs. AUC = 0.58–0.60). While other models were evaluated solely on their test sets, our model's AUC of 0.751 (95% CI: 0.692–0.820) surpassed Wang et al.'s model (AUC = 0.742) but was marginally lower than Sunaga et al.'s model (AUC = 0.78). However, test set performance alone may lack stability, as data distribution can affect model outcomes. Notably, our hyperparameters were determined based on 5-fold cross-validation of the training set rather than the test set, as was done in some other studies. Although this approach may reduce the AUC in the test set, it helps prevent overfitting and provides a more reliable estimate of the model's performance on unseen data. External validation with two independent cohorts further demonstrated the robustness of our model.

Recent research highlights the importance of composited indicators in predicting IVIG resistance,[21–25] and in line with these findings, our model incorporated several such indicators, including CAR, PNI, NLR, and SII. Notably, SHAP analysis consistently ranked CAR and PNI among the top three features, further supporting their critical role. Additionally, sex was consistently ranked among the top three features, a finding aligned with previous randomized controlled trials in North America.[26] Considering that sex is also an important demographic indicator, we conducted a stratified analysis based on sex to further analyze its effect on other predictor variables and validate the reliability of the ranking of predictor variables for IVIG resistance. The results suggest that CAR and PNI were the most effective predictors for IVIG resistance in KD patients regardless of sex. As KD is a form of vasculitis, the role of inflammatory biomarkers in predicting IVIG resistance and coronary artery lesions (CALs) has garnered significant attention. Previous studies have identified traditional inflammatory markers such as CRP, NEU%, WBC, and ALB as independent predictors of IVIG resistance.[27–29] In our model, absolute lymphocyte count (ALC) and CRP consistently ranked between 4th and 7th in SHAP plots, underscoring their stable contribution to predictive accuracy. As an important demographic indicator, no clear guideline currently specifies which age group of children is more susceptible to IVIG resistance. Some studies suggest that KD patients under 12 months of age are more likely to develop CALs.[30] Therefore, in our study, age is treated as a binary variable with

Table 2. Performance of six models in the test set and two external validation sets

Models	Wenzhou Hospital test set				Shaoxing Hospital				Quzhou Hospital			
	AUC[a]	Sensitivity[a]	Specificity[a]	Accuracy[a]	AUC[a]	Sensitivity[a]	Specificity[a]	Accuracy[a]	AUC[a]	Sensitivity[a]	Specificity[a]	Accuracy[a]
Random forest	0.751 (0.692–0.820)	0.726 (0.631–0.821)	0.720 (0.681–0.764)	0.721 (0.686–0.758)	0.827 (0.659–0.949)	0.800 (0.400–1.000)	0.863 (0.800–0.926)	0.860 (0.790–0.920)	0.735 (0.595–0.860)	1.000 (1.000–1.000)	0.613 (0.500–0.713)	0.635 (0.529–0.729)
XGBoost	0.729 (0.667–0.801)	0.631 (0.524–0.738)	0.762 (0.724–0.799)	0.743 (0.708–0.780)	0.756 (0.644–0.878)	1.000 (1.000–1.000)	0.663 (0.568–0.747)	0.680 (0.590–0.760)	0.677 (0.470–0.900)	1.000 (1.000–1.000)	0.375 (0.275–0.475)	0.412 (0.318–0.506)
Logistic regression	0.725 (0.664–0.787)	0.738 (0.643–0.833)	0.640 (0.598–0.681)	0.654 (0.618–0.693)	0.827 (0.726–0.918)	1.000 (1.000–1.000)	0.716 (0.621–0.811)	0.730 (0.640–0.820)	0.577 (0.300–0.820)	0.800 (0.400–1.000)	0.425 (0.313–0.538)	0.447 (0.341–0.553)
LGBM	0.738 (0.671–0.800)	0.488 (0.381–0.595)	0.880 (0.850–0.907)	0.824 (0.791–0.853)	0.893 (0.808–0.968)	1.000 (1.000–1.000)	0.811 (0.726–0.884)	0.820 (0.740–0.890)	0.743 (0.540–0.888)	0.800 (0.400–1.000)	0.738 (0.650–0.825)	0.741 (0.647–0.835)
SVM	0.659 (0.593–0.727)	0.702 (0.605–0.809)	0.620 (0.582–0.659)	0.632 (0.595–0.666)	0.808 (0.599–0.954)	0.800 (0.333–1.000)	0.832 (0.599–0.954)	0.830 (0.750–0.900)	0.600 (0.200–0.915)	0.400 (0.000–1.000)	0.925 (0.863–0.975)	0.894 (0.824–0.953)
CatBoost	0.733 (0.668–0.807)	0.595 (0.500–0.702)	0.811 (0.774–0.848)	0.780 (0.747–0.814)	0.785 (0.627–0.924)	0.800 (0.400–1.000)	0.747 (0.663–0.842)	0.750 (0.670–0.840)	0.633 (0.435–0.833)	1.000 (1.000–1.000)	0.375 (0.275–0.475)	0.412 (0.318–0.506)

AUC, area under the receiver operating characteristic curve; XGBoost, eXtreme gradient boosting; LGBM, light gradient boosting machine; SVM, support vector machine; CatBoost, categorical boosting.

[a]Reported 95% confidence intervals (in brackets) are obtained via percentile distributions of 1000-fold bootstrap resampling on the validation partition.

12 months as the cutoff. Age is positioned at a moderate level in the SHAP ranking. Previous studies have demonstrated that Sodium (NA),[11,31] aspartate aminotransferase (AST),[11,12,31] alanine aminotransferase (ALT),[10,31] and total bilirubin (TBIL)[12,31] are closely associated with IVIG resistance. These variables exert a moderate predictive influence on the model, similar to age. However, markers such as WBC and ANC showed limited impact, suggesting that their predictive significance in IVIG resistance may have been overestimated.

It is important to note that while we developed several machine learning models using various algorithms, SHAP analysis was conducted exclusively on the random forest model. Despite its slightly lower performance in external validation compared to the LGBM model, the random forest model demonstrated robust and consistent results across both internal and external evaluations. Conversely, although the LGBM model excelled in external validation, it performed poorly in 5-fold cross-validation. Given that the test set used for 5-fold cross-validation was significantly larger than the external validation cohorts, we prioritized the random forest model for SHAP analysis and comparison with previous models. In terms of variable selection, the lack of established guidelines for feature inclusion posed challenges in determining the optimal number of features. While more features can enrich a model, excessive inclusion may hinder its clinical applicability, and non-causal features could reduce predictive accuracy. To enhance reliability, we focused on objective indicators, such as sex, age, and laboratory results, which were drawn from prior research or established clinical practice, ensuring a strong causal link to outcomes. Notably, all variables included in the model are readily available in routine clinical practice, enhancing its practicality and ease of implementation in diverse clinical settings.

Limitations of the study

This study has limitations. First, the size of the external validation set is relatively small, and further validation in larger, independent cohorts is necessary to confirm the model's effectiveness in broader populations. Second, prospective studies are needed to fully validate the performance of the random forest model.

Conclusions

The primary findings of this study are as follows. (1) We successfully developed an effective machine learning-based model to predict IVIG resistance in KD patients. (2) CAR, PNI, and sex were identified as key predictors of IVIG resistance in these patients. We anticipate that our model could be integrated into electronic health record systems to provide clinical decision support in the near future. However, incorporating additional data, such as genetic variants, could further enhance the predictive power of the model.

RESOURCE AVAILABILITY

Lead contact

Further requests for resources and materials should be directed to and will be fulfilled by the lead contact, Xiu-Feng Huang (hxfwzmc@163.com).

Materials availability

No materials were used or generated in this study.

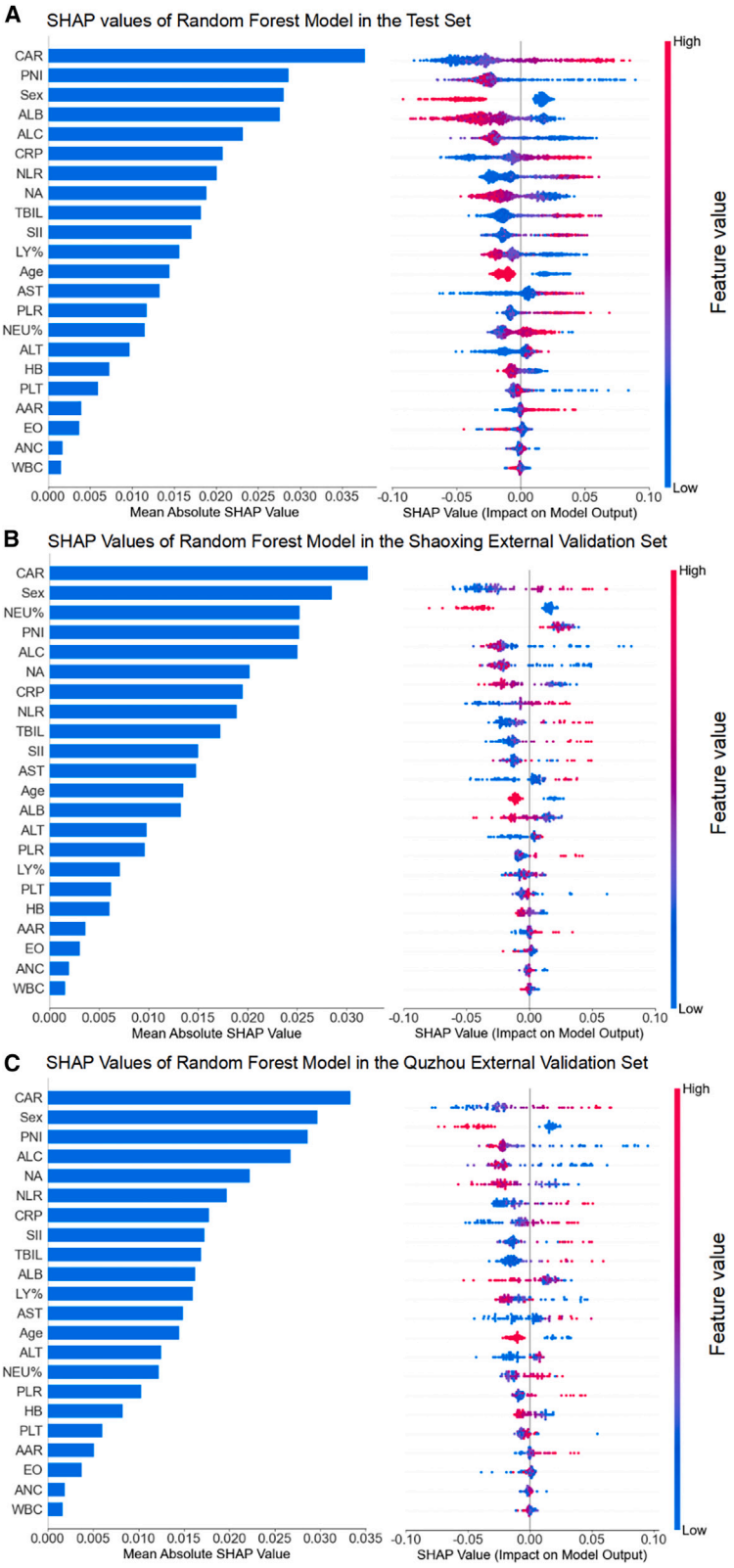


Figure 3. The 20 most important features for predicting IVIG resistance of random forest model

(A–C) The top 20 features for the random forest model were ranked. Blue indicates low values (or 0 for binary features that are absent), while red represents high values (or 1 for binary features that are present). Sex and age are represented as binary variables. Male is coded as 0 and female as 1. Age is coded as 0 for patients aged 12 months or younger and 1 for those older than 12 months.

Table 3. Comparison of previous scoring systems and our model in the test set

Models	Study population	AUC	Sensitivity	Specificity	Accuracy
Egami	Japan	0.650	0.488	0.811	0.765
Kobayashi	Japan	0.646	0.476	0.815	0.767
Sano	Japan	0.658	0.560	0.756	0.728
Formosa	Taiwan, China	0.590	0.750	0.429	0.475
Fu	Beijing, China	0.621	0.452	0.789	0.742
Random forest	Wenzhou, China	0.751	0.726	0.720	0.721

AUC, area under the receiver operating characteristic curve.

Data and code availability

The clinical data reported in this study cannot be deposited in a public repository because of ethical constraints. To request access to the data from the machine learning set, contact the lead contact, Xiu-Feng Huang (hxfwzmc@163.com). The code reported in this paper will be shared by the lead contact upon request. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (82370509) and Natural Science Foundation of Zhejiang Province (LY23H120002).

AUTHOR CONTRIBUTIONS

Y.X. and Y.H. had full access to all of the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. All authors reviewed the manuscript and agreed to be accountable for all aspects of the work, ensuring the accuracy and integrity of the publication. Conceptualization, X.-F.H. and M.C.; data curation, M.G., W.L., Y.M., H.Q., H.W., Hui Zhang, Hao Zhang, and L.W.; writing – original draft, Y.X. and Y.H.; writing – review and editing, X.-F.H., M.C., and X.-L.C.; visualization, Y.X. and Y.H.; funding acquisition, M.C. and X.-F.H.; project administration, R.W. and X.R.; supervision, M.C. and X.-F.H.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Data collection
  - Definition
  - Model construction and validation
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.112004.

Received: October 21, 2024
Revised: December 20, 2024
Accepted: February 7, 2025
Published: February 11, 2025

REFERENCES

1. Kawasaki, T. (1967). [Acute febrile mucocutaneous syndrome with lymphoid involvement with specific desquamation of the fingers and toes in children]. Arerugi *16*, 178–222.
2. Newburger, J.W., Takahashi, M., and Burns, J.C. (2016). Kawasaki Disease. J. Am. Coll. Cardiol. *67*, 1738–1749. https://doi.org/10.1016/j.jacc.2015.12.073.
3. McCrindle, B.W., Rowley, A.H., Newburger, J.W., Burns, J.C., Bolger, A.F., Gewitz, M., Baker, A.L., Jackson, M.A., Takahashi, M., Shah, P.B., et al. (2017). Diagnosis, Treatment, and Long-Term Management of Kawasaki Disease: A Scientific Statement for Health Professionals From the American Heart Association. Circulation *135*, e927–e999. https://doi.org/10.1161/cir.0000000000000484.
4. Nakamura, Y., Yashiro, M., Uehara, R., Oki, I., Watanabe, M., and Yanagawa, H. (2008). Epidemiologic features of Kawasaki disease in Japan: results from the nationwide survey in 2005-2006. J. Epidemiol. *18*, 167–172. https://doi.org/10.2188/jea.je2008001.
5. Son, M.B.F., Gauvreau, K., Ma, L., Baker, A.L., Sundel, R.P., Fulton, D.R., and Newburger, J.W. (2009). Treatment of Kawasaki disease: analysis of 27 US pediatric hospitals from 2001 to 2006. Pediatrics *124*, 1–8. https://doi.org/10.1542/peds.2008-0730.
6. Burns, J.C., Capparelli, E.V., Brown, J.A., Newburger, J.W., and Glode, M.P. (1998). Intravenous gamma-globulin treatment and retreatment in Kawasaki disease. US/Canadian Kawasaki Syndrome Study Group. Pediatr. Infect. Dis. J. *17*, 1144–1148. https://doi.org/10.1097/00006454-199812000-00009.
7. Uehara, R., Belay, E.D., Maddox, R.A., Holman, R.C., Nakamura, Y., Yashiro, M., Oki, I., Ogino, H., Schonberger, L.B., and Yanagawa, H. (2008). Analysis of potential risk factors associated with nonresponse to initial intravenous immunoglobulin treatment among Kawasaki disease patients in Japan. Pediatr. Infect. Dis. J. *27*, 155–160. https://doi.org/10.1097/INF.0b013e31815922b5.
8. Chbeir, D., Gaschignard, J., Bonnefoy, R., Beyler, C., Melki, I., Faye, A., and Meinzer, U. (2018). Kawasaki disease: abnormal initial echocardiogram is associated with resistance to IV Ig and development of coronary artery lesions. Pediatr. Rheumatol. Online J. *16*, 48. https://doi.org/10.1186/s12969-018-0264-7.
9. Chen, S., Dong, Y., Yin, Y., and Krucoff, M.W. (2013). Intravenous immunoglobulin plus corticosteroid to prevent coronary artery abnormalities in Kawasaki disease: a meta-analysis. Heart *99*, 76–82. https://doi.org/10.1136/heartjnl-2012-302126.
10. Egami, K., Muta, H., Ishii, M., Suda, K., Sugahara, Y., Iemura, M., and Matsuishi, T. (2006). Prediction of resistance to intravenous immunoglobulin treatment in patients with Kawasaki disease. J. Pediatr. *149*, 237–240. https://doi.org/10.1016/j.jpeds.2006.03.050.
11. Kobayashi, T., Inoue, Y., Takeuchi, K., Okada, Y., Tamura, K., Tomomasa, T., Kobayashi, T., and Morikawa, A. (2006). Prediction of intravenous immunoglobulin unresponsiveness in patients with Kawasaki disease.

Circulation *113*, 2606–2612. https://doi.org/10.1161/circulationaha.105.592865.

12. Sano, T., Kurotobi, S., Matsuzaki, K., Yamamoto, T., Maki, I., Miki, K., Kogaki, S., and Hara, J. (2007). Prediction of non-responsiveness to standard high-dose gamma-globulin therapy in patients with acute Kawasaki disease before starting initial treatment. Eur. J. Pediatr. *166*, 131–137. https://doi.org/10.1007/s00431-006-0223-z.
13. Lin, M.T., Chang, C.H., Sun, L.C., Liu, H.M., Chang, H.W., Chen, C.A., Chiu, S.N., Lu, C.W., Chang, L.Y., Wang, J.K., and Wu, M.H. (2016). Risk factors and derived formosa score for intravenous immunoglobulin unresponsiveness in Taiwanese children with Kawasaki disease. Journal of the Formosan Medical Association = Taiwan yi zhi *115*, 350–355. https://doi.org/10.1016/j.jfma.2015.03.012.
14. Yang, S., Song, R., Zhang, J., Li, X., and Li, C. (2019). Predictive tool for intravenous immunoglobulin resistance of Kawasaki disease in Beijing. Arch. Dis. Child. *104*, 262–267. https://doi.org/10.1136/archdischild-2017-314512.
15. Liu, J., Zhang, J., Huang, H., Wang, Y., Zhang, Z., Ma, Y., and He, X. (2021). A Machine Learning Model to Predict Intravenous Immunoglobulin-Resistant Kawasaki Disease Patients: A Retrospective Study Based on the Chongqing Population. Front. Pediatr. *9*, 756095. https://doi.org/10.3389/fped.2021.756095.
16. Tsai, C.M., Lin, C.H.R., Kuo, H.C., Cheng, F.J., Yu, H.R., Hung, T.C., Hung, C.S., Huang, C.M., Chu, Y.C., and Huang, Y.H. (2023). Use of Machine Learning to Differentiate Children With Kawasaki Disease From Other Febrile Children in a Pediatric Emergency Department. JAMA Netw. Open *6*, e237489. https://doi.org/10.1001/jamanetworkopen.2023.7489.
17. Lam, J.Y., Shimizu, C., Tremoulet, A.H., Bainto, E., Roberts, S.C., Sivilay, N., Gardiner, M.A., Kanegaye, J.T., Hogan, A.H., Salazar, J.C., et al. (2022). A machine-learning algorithm for diagnosis of multisystem inflammatory syndrome in children and Kawasaki disease in the USA: a retrospective model development and validation study. Lancet. Digit. Health *4*, e717–e726. https://doi.org/10.1016/s2589-7500(22)00149-2.
18. Kuniyoshi, Y., Tokutake, H., Takahashi, N., Kamura, A., Yasuda, S., and Tashiro, M. (2020). Comparison of Machine Learning Models for Prediction of Initial Intravenous Immunoglobulin Resistance in Children With Kawasaki Disease. Front. Pediatr. *8*, 570834. https://doi.org/10.3389/fped.2020.570834.
19. Wang, T., Liu, G., and Lin, H. (2020). A machine learning approach to predict intravenous immunoglobulin resistance in Kawasaki disease patients: A study based on a Southeast China population. PLoS One *15*, e0237321. https://doi.org/10.1371/journal.pone.0237321.
20. Sunaga, Y., Watanabe, A., Katsumata, N., Toda, T., Yoshizawa, M., Kono, Y., Hasebe, Y., Koizumi, K., Hoshiai, M., Kawakami, E., and Inukai, T. (2023). A simple scoring model based on machine learning predicts intravenous immunoglobulin resistance in Kawasaki disease. Clin. Rheumatol. *42*, 1351–1361. https://doi.org/10.1007/s10067-023-06502-1.
21. Liu, C., and Wu, J. (2022). Value of blood inflammatory markers for predicting intravenous immunoglobulin resistance in Kawasaki disease: A systematic review and meta-analysis. Front. Pediatr. *10*, 969502. https://doi.org/10.3389/fped.2022.969502.
22. Liu, X., Wang, L., Zhou, K., Shao, S., Hua, Y., Wu, M., Liu, L., and Wang, C. (2021). Predictive value of C-reactive protein to albumin ratio as a biomarker for initial and repeated intravenous immunoglobulin resistance in a l arge cohort of Kawasaki disease patients: a prospective cohort study. Pediatr. Rheumatol. Online J. *19*, 24. https://doi.org/10.1186/s12969-021-00517-1.
23. Chen, Y., Hua, Y., Zhang, C., Chen, S., Zhang, Q., Liao, Y., Yan, H., Wang, Y., Liu, P., Qi, J., et al. (2019). Neutrophil-to-Lymphocyte Ratio Predicts Intravenous Immunoglobulin-Resistance in Infants Under 12-Months Old With Kawasaki Disease. Front. Pediatr. *7*, 81. https://doi.org/10.3389/fped.2019.00081.
24. Li, G., Wang, T., Gou, Y., Zeng, R., Liu, D., Duan, Y., and Liu, B. (2020). Value of C-reactive protein/albumin ratio in predicting intravenous immunoglobulin-resistant Kawasaki disease- a data from multi-institutional study in China. Int. Immunopharmacol. *89*, 107037. https://doi.org/10.1016/j.intimp.2020.107037.
25. Liu, X., Shao, S., Zhang, N., Wu, M., Liu, L., Duan, H., Liu, Z., Zhou, K., Hua, Y., and Wang, C. (2022). Predictive role of sampling-time specific prognostic nutritional index cut-off values for intravenous immunoglobulin resistance and cardiovascular complications in Kawasaki disease. Int. Immunopharmacol. *110*, 108986. https://doi.org/10.1016/j.intimp.2022.108986.
26. Sleeper, L.A., Minich, L.L., McCrindle, B.M., Li, J.S., Mason, W., Colan, S.D., Atz, A.M., Printz, B.F., Baker, A., Vetter, V.L., et al. (2011). Evaluation of Kawasaki disease risk-scoring systems for intravenous immunoglobulin resistance. J. Pediatr. *158*, 831–835.e3. https://doi.org/10.1016/j.jpeds.2010.10.031.
27. Wu, S., Liao, Y., Sun, Y., Zhang, C.Y., Zhang, Q.Y., Yan, H., Qi, J.G., Liu, X.Q., Chen, Y.H., Wang, Y.L., et al. (2020). Prediction of intravenous immunoglobulin resistance in Kawasaki disease in children. World J. Pediatr. *16*, 607–613. https://doi.org/10.1007/s12519-020-00348-2.
28. Huang, H., Jiang, J., Shi, X., Qin, J., Dong, J., Xu, L., Huang, C., Liu, Y., Zheng, Y., Hou, M., et al. (2022). Nomogram to predict risk of resistance to intravenous immunoglobulin in children hospitalized with Kawasaki disease in Eastern China. Ann. Med. *54*, 442–453. https://doi.org/10.1080/07853890.2022.2031273.
29. Wang, J., Huang, X., and Guo, D. (2023). Predictors and a novel predictive model for intravascular immunoglobulin resistance in Kawasaki disease. Ital. J. Pediatr. *49*, 126. https://doi.org/10.1186/s13052-023-01531-7.
30. Manlhiot, C., Yeung, R.S.M., Clarizia, N.A., Chahal, N., and McCrindle, B.W. (2009). Kawasaki disease at the extremes of the age spectrum. Pediatrics *124*, e410–e415. https://doi.org/10.1542/peds.2009-0099.
31. Hua, W., Sun, Y., Wang, Y., Fu, S., Wang, W., Xie, C., Zhang, Y., and Gong, F. (2017). A new model to predict intravenous immunoglobin-resistant Kawasaki disease. Oncotarget *8*, 80722–80729. https://doi.org/10.18632/oncotarget.21083.
32. JCS Joint Working Group (2014). Guidelines for diagnosis and management of cardiovascular sequelae in Kawasaki disease (JCS 2013). Digest version. Circ. J. : official journal of the Japanese Circulation Society *78*, 2521–2562. https://doi.org/10.1253/circj.cj-66-0096.
33. Subspecialty Group of Cardiology the Society of Pediatrics Chinese Medical Association; Subspecialty Group of Rheumatology, the Society of Pediatrics, Chinese Medical Association; Subspecialty Group of Immunology, the Society of Pediatrics, Chinese Medical Association; Editorial Board, Chinese Journal of Pediatrics (2022). Zhonghua er ke za zhi = Chinese. J. Pediatr. *60*, 6–13. https://doi.org/10.3760/cma.j.cn112140-20211018-00879.
34. Ogata, S., Tremoulet, A.H., Sato, Y., Ueda, K., Shimizu, C., Sun, X., Jain, S., Silverstein, L., Baker, A.L., Tanaka, N., et al. (2013). Coronary artery outcomes among children with Kawasaki disease in the United States and Japan. Int. J. Cardiol. *168*, 3825–3828. https://doi.org/10.1016/j.ijcard.2013.06.027.
35. Wang, Y., Wang, W., Gong, F., Fu, S., Zhang, Q., Hu, J., Qi, Y., Xie, C., and Zhang, Y. (2013). Evaluation of intravenous immunoglobulin resistance and coronary artery lesions in relation to Th1/Th2 cytokine profiles in patients with Kawasaki disease. Arthritis Rheum. *65*, 805–814. https://doi.org/10.1002/art.37815.
36. Qiu, H., Jia, C., Wang, Z., He, Y., Rong, X., Wu, R., Chu, M., and Shi, H. (2020). Prognosis and Risk Factors of Coronary Artery Lesions before Immunoglobulin Therapy in Children with Kawasaki Disease. Balkan Med. J. *37*, 324–329. https://doi.org/10.4274/balkanmedj.galenos.2020.2020.1.56.
37. Shi, H., Weng, F., Li, C., Jin, Z., Hu, J., Chu, M., and Qiu, H. (2021). Overweight, obesity and coronary artery lesions among Kawasaki disease patients. Nutr. Metabol. Cardiovasc. Dis. *31*, 1604–1612. https://doi.org/10.1016/j.numecd.2021.01.015.
38. Breiman, L. (2001). Random forests. Mach. Learn. *45*, 199–228.

39. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Adv. Neural. Inf. Process. Syst. *30*.

40. Noble, W.S. (2006). What is a support vector machine? Nat. Biotechnol. *24*, 1565–1567.

41. Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794.

42. Hancock, J.T., and Khoshgoftaar, T.M. (2020). CatBoost for big data: an interdisciplinary review. J. Big Data *7*, 94. https://doi.org/10.1186/s40537-020-00369-8.

43. Yagin, F.H., Cicek, İ.B., Alkhateeb, A., Yagin, B., Colak, C., Azzeh, M., and Akbulut, S. (2023). Explainable artificial intelligence model for identifying COVID-19 gen e biomarkers. Comput. Biol. Med. *154*, 106619. https://doi.org/10.1016/j.compbiomed.2023.106619.

44. Shapley, L.S. (1952). A Value for n-Person Games. Contribution to the Theory of Games *2*.

45. Rodríguez-Pérez, R., and Bajorath, J. (2020). Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. J. Comput. Aided Mol. Des. *34*, 1013–1026. https://doi.org/10.1007/s10822-020-00314-0.

46. Lundberg, S., and Lee, S.I. (2017). A Unified Approach to Interpreting Model Predictions. Preprint at arXiv. https://doi.org/10.48550/arXiv.1705.07874.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Summary statistic data	This study	N/A
Software and algorithms		
Python software version (version 3.12.0)	Python Software Foundation	https://www.python.org
Scikit-learn	GitHub	https://github.com/scikit-learn/scikit-learn

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The workflow of the study is presented in Figure 1. We utilized the Kawasaki disease (KD) database from Yuying Children's Hospital, affiliated with Wenzhou Medical University (hereafter referred to as Wenzhou Hospital), to construct the predictive model. Patient data were extracted from the database for individuals admitted between January 2008 and October 2022. Inclusion criteria included: (1) meeting the diagnostic criteria for KD established by the Japanese Circulation Society (JCS)[32] (2) receiving IVIG treatment at a dose of 2g/kg during the acute treatment period, and (3) availability of comprehensive clinical data and not having extreme biomarker values (i.e., considered erroneous).

The studies involving human participants were reviewed and approved by the local ethics committee of Shaoxing Hospital, Quzhou Hospital, and the Second Affiliated Hospital of Wenzhou Medical University and Yuying Children's Hospital. Written informed consent from the participant's legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and institutional requirements.

METHOD DETAILS

Data collection

We collected two demographic characteristics (sex and age), fourteen laboratory variables, and six composite variables from electronic medical records as features for the models (Table S6). These features were selected based on clinical expertise and findings from previous studies. All laboratory variables were measured prior to IVIG infusion and included: white blood cell count (WBC), absolute neutrophil count (ANC), absolute lymphocyte count (ALC), percentage of neutrophils (NEU%), percentage of lymphocytes (LY %), eosinophil count (EO), platelet count (PLT), hemoglobin level (HB), C-reactive protein level (CRP), serum albumin level (ALB), alanine aminotransferase level (ALT), aspartate aminotransferase level (AST), total bilirubin level (TBIL), and serum sodium level (NA). Composite variables included the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), C-reactive protein-to-albumin ratio (CAR), aspartate aminotransferase-to-alanine aminotransferase ratio (AAR), systemic immune-inflammatory index (SII), and prognostic nutritional index (PNI), derived from the aforementioned laboratory variables. Demographic characteristics such as sex and age (categorized at 12 months) were treated as categorical variables. Additionally, medical notes detailing body temperature, symptoms, and treatments were reviewed to confirm the diagnosis of KD and IVIG resistance.

Definition

Complete KD was defined as having the presence of at least four principal symptoms alongside persistent fever ($\geq$5 days), and incomplete KD as having three principal manifestations in addition to persistent fever or two principal symptoms in addition to persistent fever and CALs.[32] IVIG-resistant was defined as KD patients with a persistent or recurrence of fever $\geq$38°C at any time from 36 h to 2 weeks after initial IVIG treatment accompanied by one or more of the main symptoms.[33] Coronary artery diameter exhibits variations among different races.[34] Thus, diagnostic criteria for CALs, tailored specifically for Chinese children, were adopted[35–37]: coronary artery diameter $\geq$2.5 mm in patients aged 0-3 years old, $\geq$3.0 mm in patients aged 3-9 years old, and $\geq$3.5 mm in patients aged older than 9 years.

Model construction and validation

The study design is shown in Figure 1. We used time series splitting to divide the machine learning set into the training set for constructing models and the test set for internal validation at a 7:3 ratio. Six machine learning algorithms, including the logistic regression analysis with L2 regularization, random forest,[38] Light Gradient Boosting Machine (LGBM),[39] Support Vector Machine (SVM),[40] eXtreme Gradient Boosting (XGBoost),[41] and Categorical Boosting (CatBoost)[42] were employed in the training set to construct the model. The 5-fold cross-validation was used to optimize the hyperparameters of each model. The training set was split into 5 smaller sets. The model was trained using 4 sets (80% of the original dataset) and validated using the remaining set (20% of the

original dataset). This process was repeated 5 times, with each of the 5 sets serving as the validation set exactly once. The final performance measure reported by the cross-validation is the average of the values computed in all 5 folds. The hyperparameter that maximizes the average area under the receiver operating characteristic (ROC) curve (AUC) of 5-fold cross-validation was identified as the optimal hyperparameter. The maximized AUC of the models was used to evaluate the performance of each model. The classification threshold corresponding to the maximum Youden index represents the optimal threshold. The sensitivity, specificity, and accuracy of the models are determined based on the optimal threshold. The adequacy of the sample size for model construction was evaluated using the learning curve. Keeping the optimal hyperparameters, the test set and two external validation sets were employed to verify the models' generalization ability.

Given the severe imbalance between IVIG-responsive and IVIG-resistant patients, we employed the Support Vector Machine Synthetic Minority Over-sampling Technique(SVM-SMOTE),[43] which enhances the effectiveness of oversampling in improving the performance of machine learning models compared to SMOTE. The SHapley Additive exPlanations (SHAP) was a unified approach for explaining the outcome of the machine learning model and was used to quantify and rank the importance of features.[44–46] SHAP values assess the significance of model output, where a higher SHAP value signifies a feature's heightened impact and greater importance within the model.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed using and Python software (version 3.12.0). Measurement data were expressed as medians (P25-P75), while categorical data were represented as percentages. The Chi-square test or Fisher's exact test was employed for categorical variables, the independent two-sample t-test for normally distributed numerical variables, and the non-parametric Mann-Whitney U test for non-normally distributed data. All tests were two-tailed, with a significance threshold of $P < 0.05$. Accuracy, sensitivity, specificity, and AUC were used to evaluate model performance. For 5-fold cross-validation, results were expressed as means and ranges. For test set validation and external validations, results were reported with 95% confidence intervals, obtained via percentile distributions from 1000-fold bootstrap resampling of the validation partition. The DeLong test was used to compare AUCs across different models.